Nico Kreiling

Wie Generative KI eine Personalisierungs-Revolution auslöst

#ittage

scieneers

DRIVEN BY DATA

We gain knowledge from data and create value. For our customers, society and ourselves.

Nico Kreiling

Data Scientist @ scieneers

Host of techtiefen.de

Personalization

“Adapt content as precisely as possible to the preferences and interests of the addressee”

Wähle einen Konferenztag aus:

Dienstag, 12. Dez 2023

Was sind Deine Interessen? Was möchtest Du Dir auf den IT-Tagen anhören?

Programmplan erstellen!

scieneers
DRIVEN BY DATA

🤖 Plan-O-Mat: Dein individueller Planner für die IT-Tage 2023

⚠️ Chatverlauf löschen!

Hallo! 👋

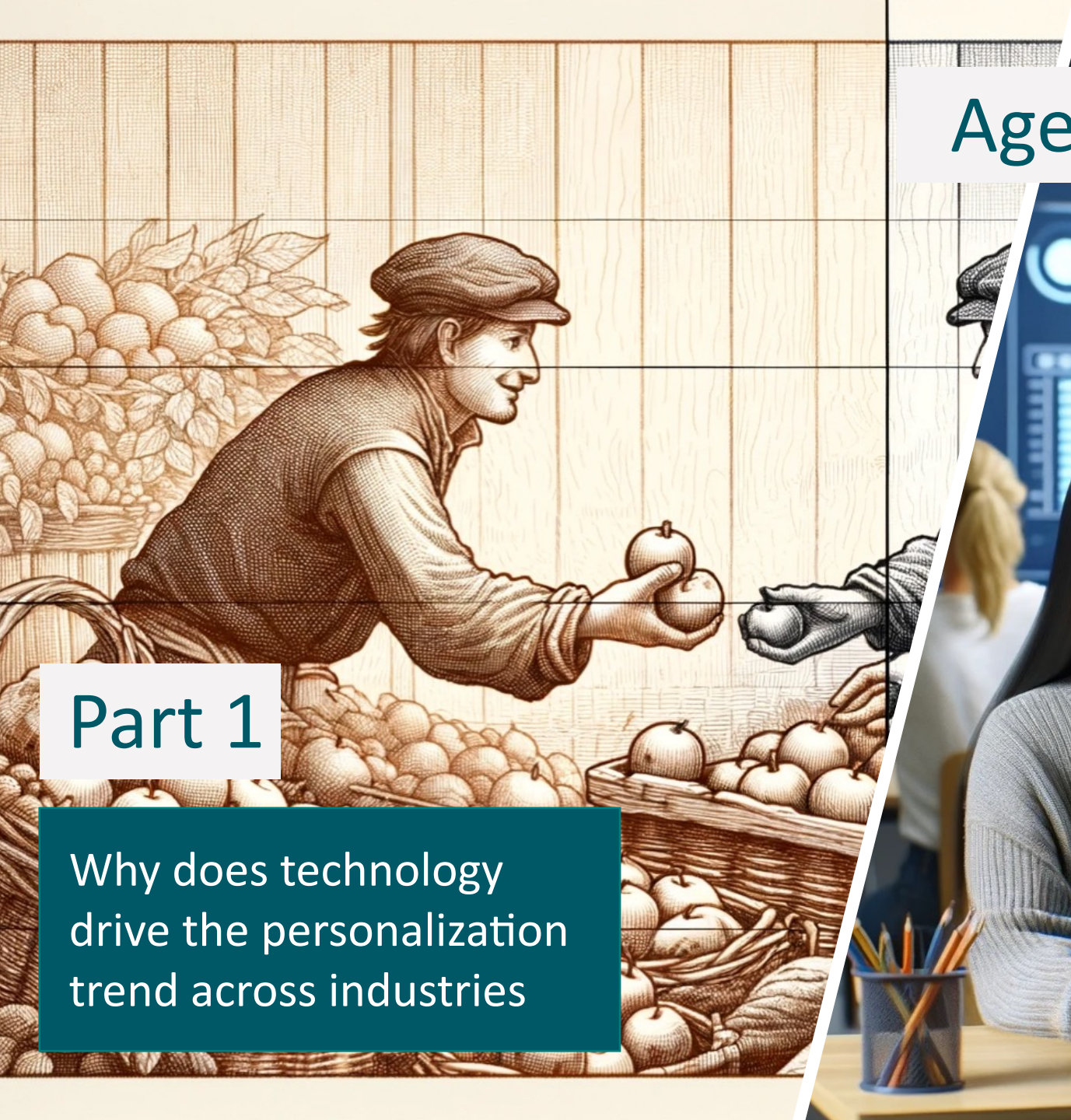
Ich bin Plan-O-Mat, Dein individueller AI-Assistent zur Planung Deines Programms auf den IT-Tage 2023. **Wähle** dazu als erstes in der Sidebar einen **Konferenztag aus**. **Formuliere** anschließend Deine **Interessen**. 💬

Eine Eingabe kann beispielsweise folgendermaßen aussehen:

> *Hi, ich bin Anja, 30 Jahre alt und arbeite seit drei Jahren als Data Scientist bei datapath. Bei meiner Arbeit programmiere ich viele KI-Anwendungen mit Python. Die Themen "KI" und "Machine Learning" interessieren mich sehr. Die IT-Tage möchte ich nutzen, um mich über Neuigkeiten in diesen Bereichen zu informieren und weiterzubilden. Ich bevorzuge technisch anspruchsvolle Vorträge und Vorstellung von konkreten Anwendungsfällen.*

Ich erstelle Dir dann Deinen individuellen Programmplan anhand Deiner Interessen.

Agenda



Part 1

Why does technology drive the personalization trend across industries

Part 2

How is AI used in real-world project right now?

12.000 b.c.

Direct Trade

# of products: < 50
Restrictions: same place and time

10. Jahrhundert

Grocery stores

# of products: < 100
Restrictions: specialized stores, multiple stops required

1930

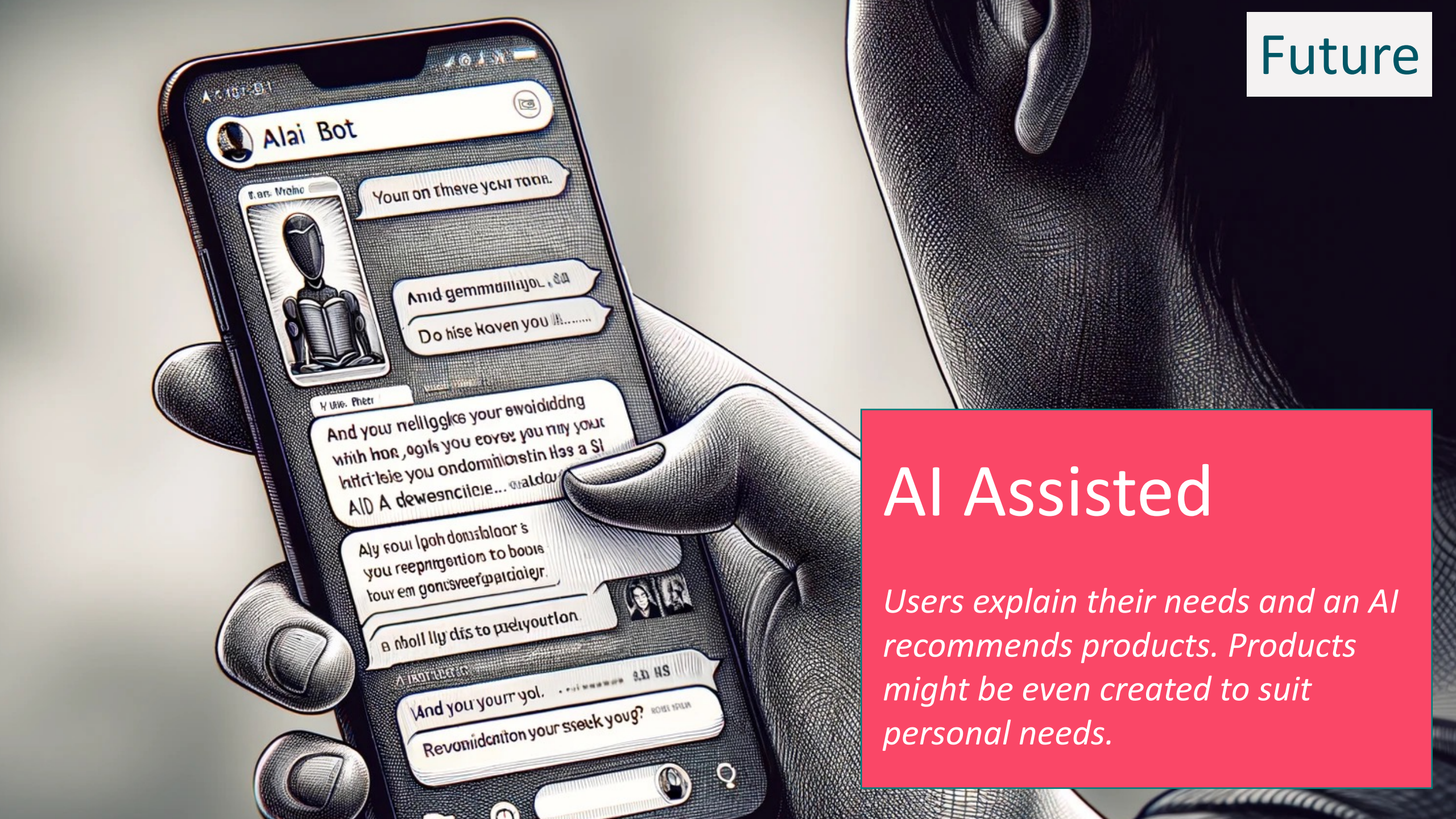
Supermarket

# of products: thousands
Restrictions: limited storage space

2000

Online Shopping

# of producs: millions
Limitation: Existing products, Need to find suitable things



Future

AI Assisted

Users explain their needs and an AI recommends products. Products might be even created to suit personal needs.

AI Generated Childrensbooks

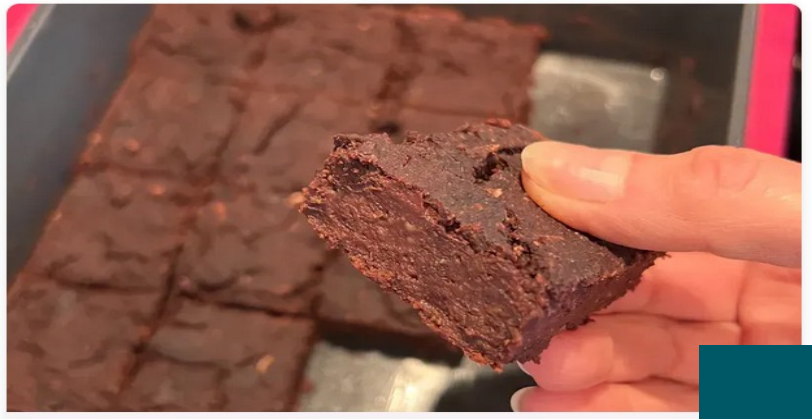
https://childbook.ai

All Recipes

FILTERED RESULTS

Roasted Red Pepper and Potato Soup

04/11/2023 ElleBHessler



Decadent Date Black Bean Brownies

02/11/2023 PG

AI Generated Recipes

https://www.dishgen.com/

40.000 b.c.

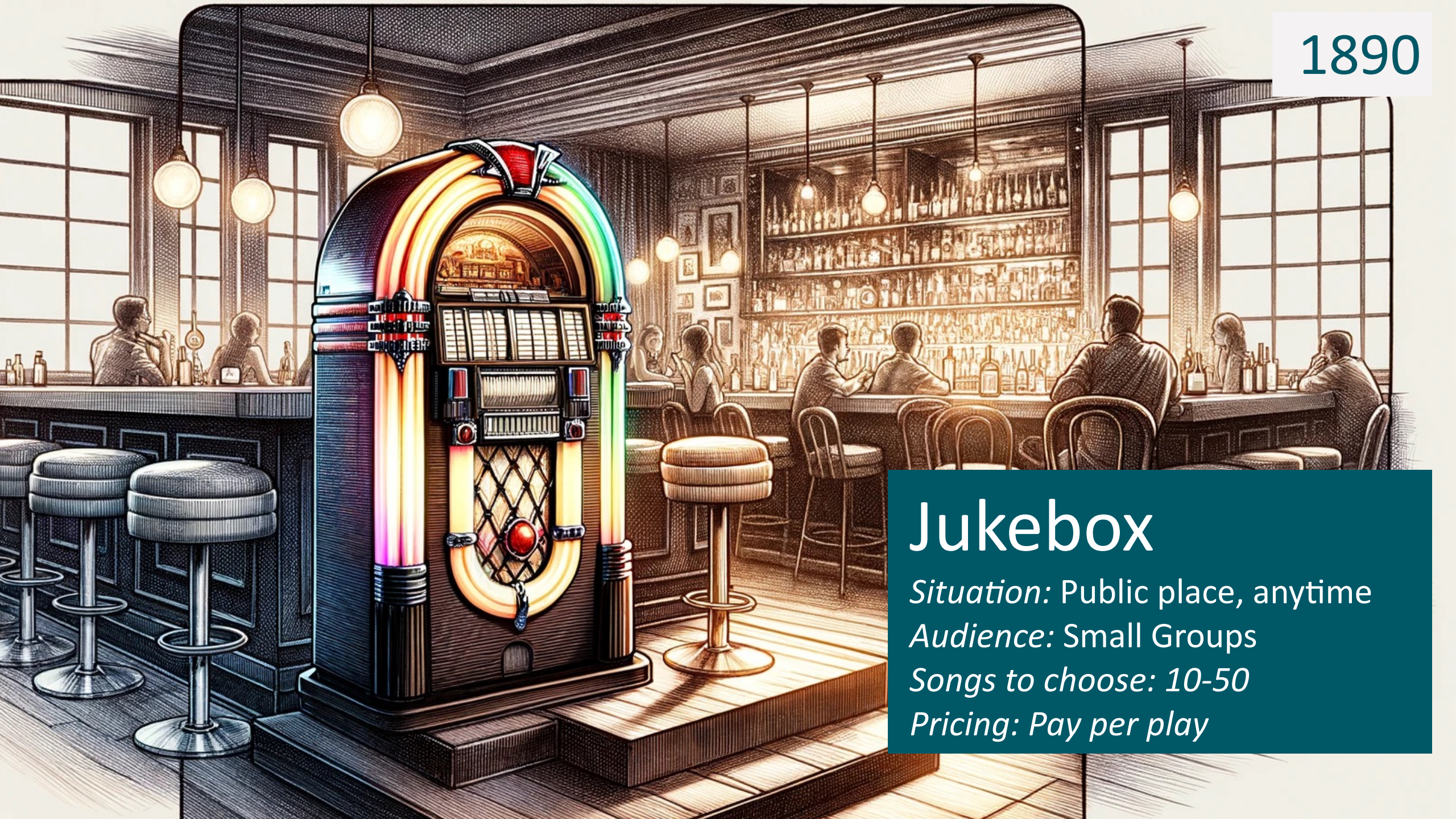
Live Music

Situation: One place, one time

Audience: Small-Medium Groups

Songs to choose: None

Pricing: Tickets / Invites



1890

Jukebox

Situation: Public place, anytime

Audience: Small Groups

Songs to choose: 10-50

Pricing: Pay per play

20th century

Vinyl Records

Situation: At home, anytime

Audience: Family members

Songs to choose: thousands

Pricing: But records

1980s

Walkman

Situation: anywhere, anytime

Audience: individual

Songs to choose: few millions

Pricing: Buy tapes

2006

Spotify

Situation: anywhere, anytime

Audience: individual

Songs to choose: > 100 millions

Pricing: Subscription

Future

AI Generate Music

Situation: anywhere, anytime

Audience: individual

Songs to choose: infinite

Pricing: ?

scieneers
DRIVEN BY DATA

Artificial Music

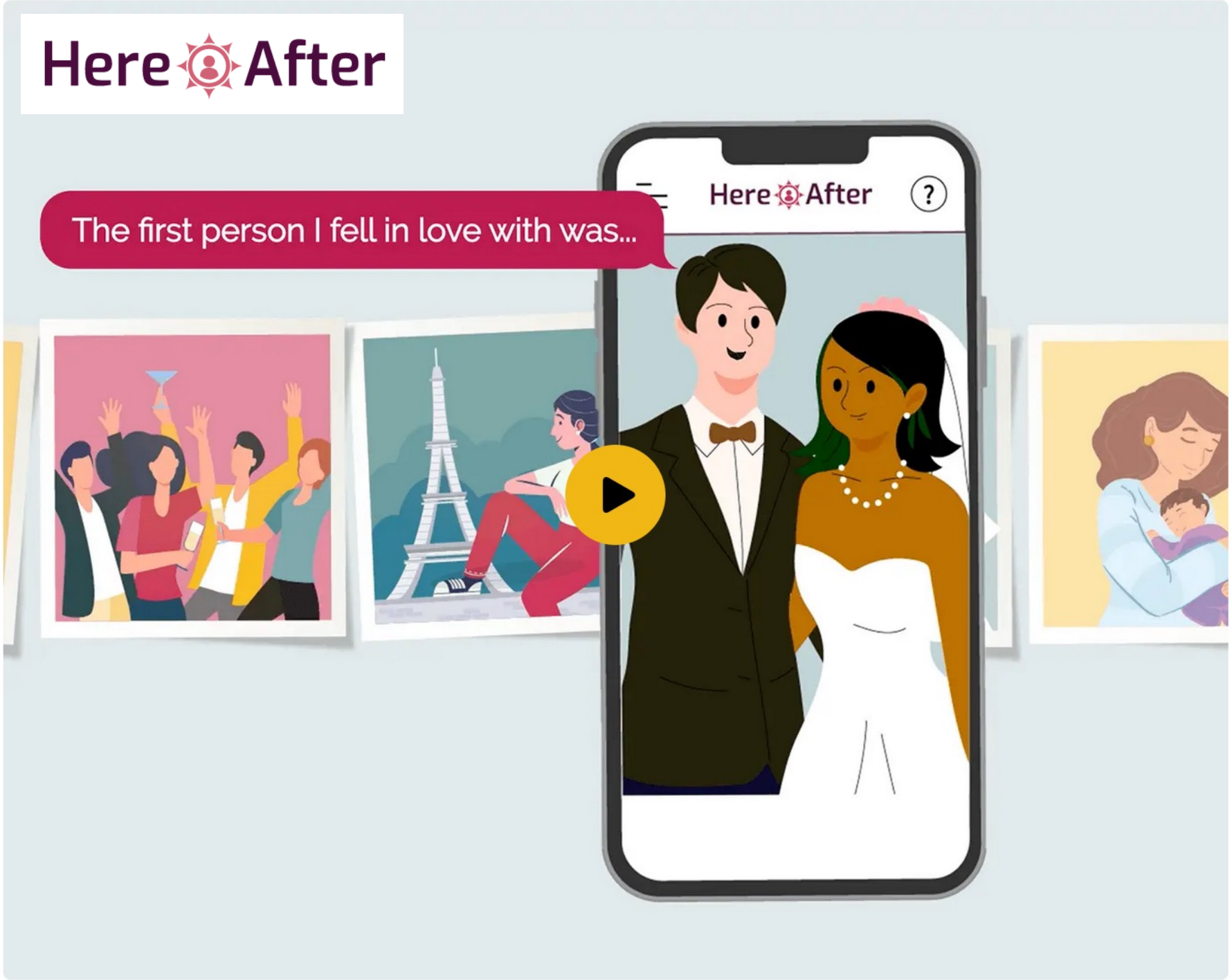
The song “Heart on my sleeve” uses the voice of Drake & The Weeknd and become viral on TicToc and YouTube

Artists **Grimes** willing to 'split 50% royalties' with AI-generated music

scieneers
DRIVEN BY DATA

AI Translated Podcasts

Spotify is working on a project to use podcasters original voice speaking other languages



scieneers
DRIVEN BY DATA

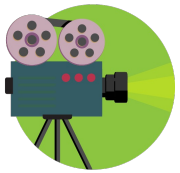
Interaction with Ancestors

Answer a couple of questions and create an interactive Avatar for the infinity

What to expect from AI

The news, that are most relevant to you, customized in length and complexity for your needs.

Always new songs that fit your mood and taste. AI generated podcasts by your social media feed.



Much more high-quality video-animation movies on niche topics (Youtube).

True open-world games, more intelligent interaction, Metaverse?

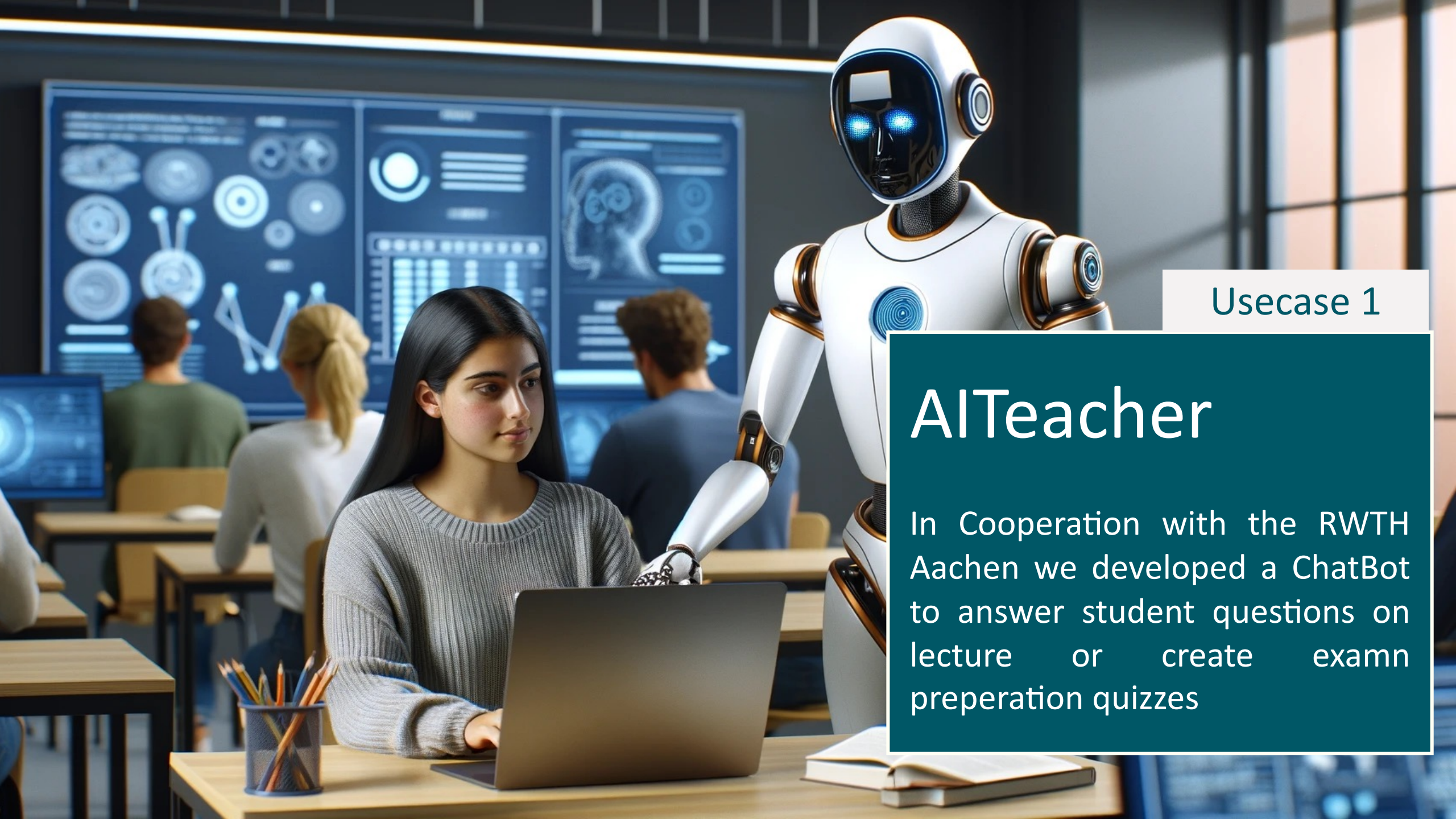
More High-Quality-Visualizations. Interactive-Videos created from images.

Real-World Use-Cases we work(ed) on...

AITeacher

Multi-Agent-Systems

Personalization



Usecase 1

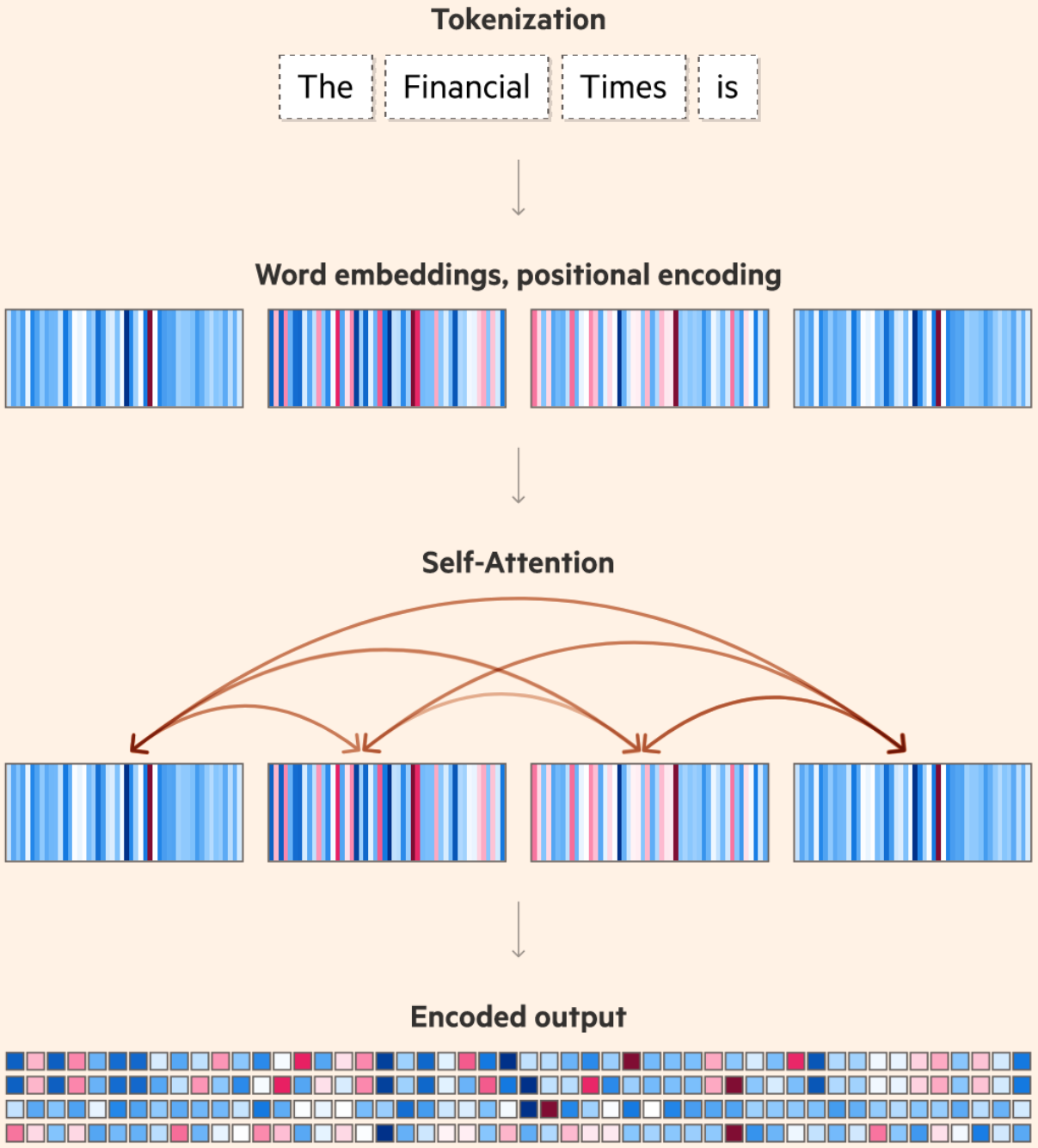
AITeacher

In Cooperation with the RWTH Aachen we developed a ChatBot to answer student questions on lecture or create examn preperation quizzes

Most important things

- If you want to understand how embeddings and attention works, read this article: https://ig.ft.com/generative-ai/

- Context-Size: How many tokens can the model handle?



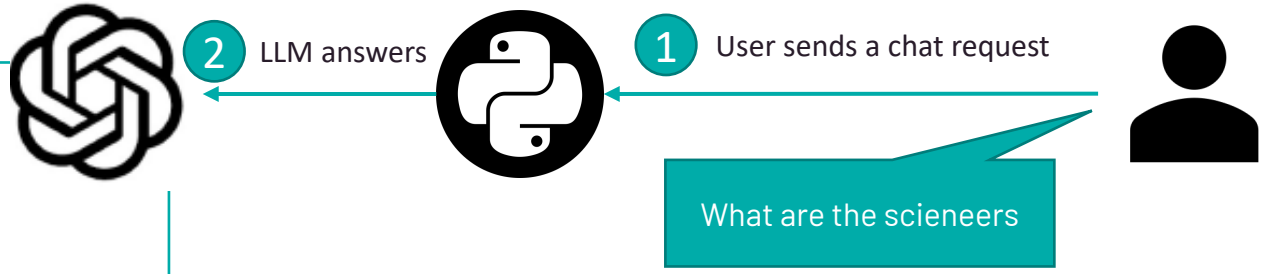
SOURCE: https://ig.ft.com/generative-ai/

Retrieval Augmented Generation (RAG)

- Use some search (mostly vector-based) to identify relevant information
- Prompt the model to answer the user question by using those information
- Major Benefits:
  - No model training
  - Easy to update knowledge information
  - Linking of sources

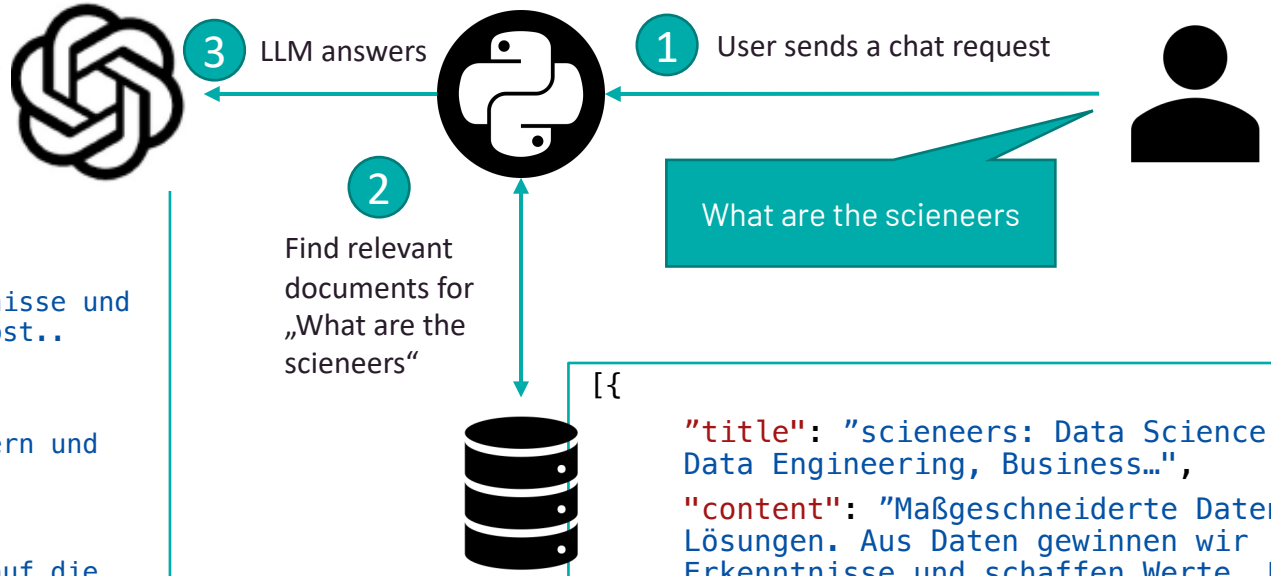
How do classical LLM Calls look like

```
[{
    "role": "system",
    "content": "You are an AI assistant that
    helps people find information."
},{
    "role": "user",
    "content": ”What are the scieneers?"
},{
    "role": "assistant",
    "content": "I'm sorry, but I am not familiar
    with the term \"scieneers\". Can you please
    provide more context or information about
    what you are referring to? "
}]
```



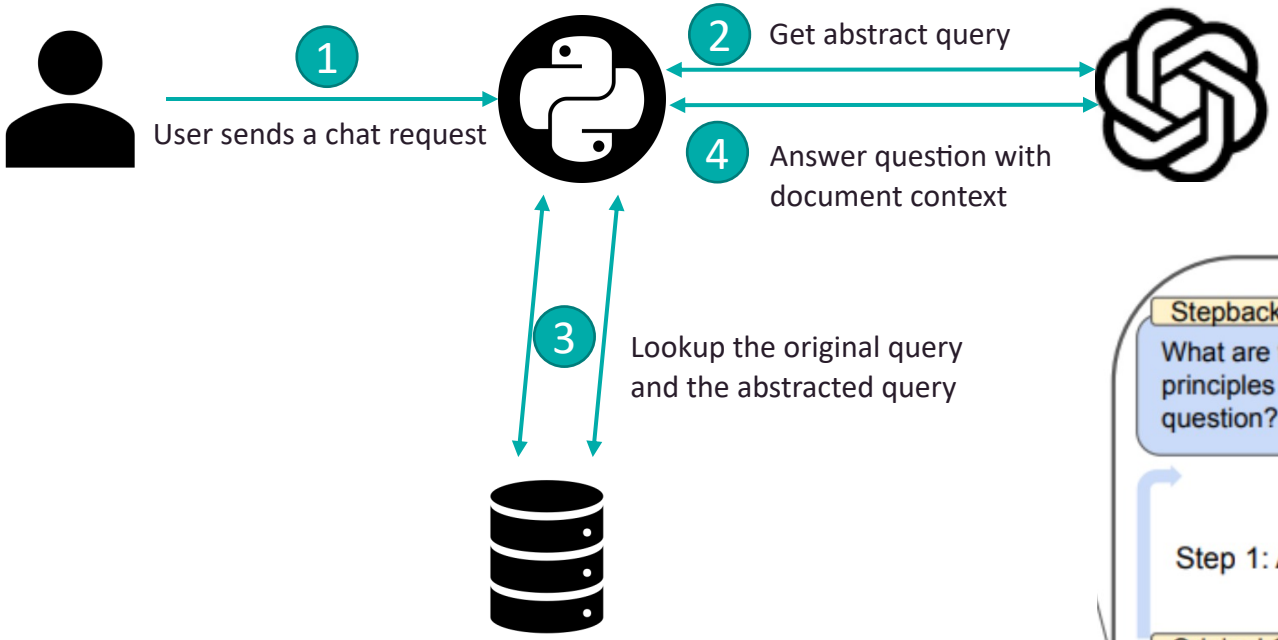
How do retriever enhanced LLM (RAG) Calls look like

```
[{
    "role": "system", "content": "You are an AI assistant that helps find information."
},{
    "role": "user", "content": ”What are the scieneers? Nutze zur beantwortung folgenden Context:
    - scieneers: Data Science, Data Engineering, Business…
    Maßgeschneiderte Daten-Lösungen. Aus Daten gewinnen wir Erkenntnisse und schaffen Werte. Für unsere Kunden, die Gesellschaft und uns selbst..
    - Scieneers als Arbeitgeber
    Ist scieneers GmbH der richtige Arbeitgeber für Dich? Anonyme Erfahrungsberichte zu Gehalt, Kultur und Karriere von Mitarbeitern und Bewerbern finde”
},{
    "role": "assistant", "content": "Die "scieneers" beziehen sich auf die scieneers GmbH, ein Unternehmen, das sich auf Datenwissenschaft (Data Science), Datenengineering (Data Engineering) und Geschäftslösungen (Business Solutions) spezialisiert hat. Ihr Fokus liegt auf der Bereitstellung maßgeschneiderter Datenlösungen, um aus Daten Erkenntnisse zu gewinnen und Werte zu schaffen. Dies beinhaltet sowohl die Unterstützung von Kunden als auch die Förderung der Gesellschaft und des eigenen Unternehmens.
    Als Arbeitgeber bietet die scieneers GmbH potenziellen Mitarbeitern und Bewerbern Karrieremöglichkeiten in diesen Bereichen. Die Bewertungen von Mitarbeitern und Bewerbern zu Gehalt, Unternehmenskultur und Karriereentwicklung können variieren, daher ist es ratsam, anonyme Erfahrungsberichte zu recherchieren, um ein umfassendes Bild des Unternehmens als Arbeitgeber zu erhalten."
}]
```

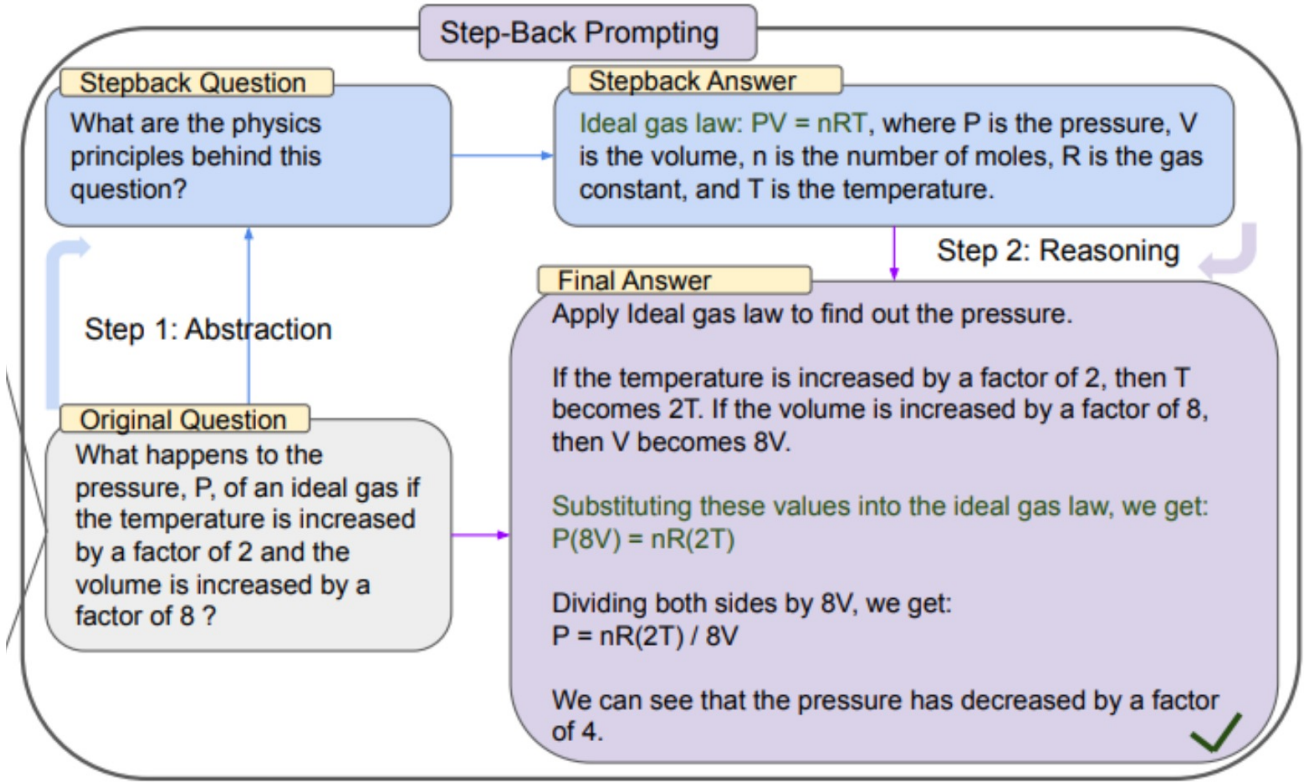



```
[{
    ”title": ”scieneers: Data Science, Data Engineering, Business…",
    "content": ”Maßgeschneiderte Daten-Lösungen. Aus Daten gewinnen wir Erkenntnisse und schaffen Werte. Für unsere Kunden, die Gesellschaft und uns selbst.."
},{
    ”title": ”Scieneers als Arbeitgeber",
    "content": ”Ist scieneers GmbH der richtige Arbeitgeber für Dich? Anonyme Erfahrungsberichte zu Gehalt, Kultur und Karriere von Mitarbeitern und Bewerbern finde"
}]
```

Retrieval enhance LLM with query manipulation



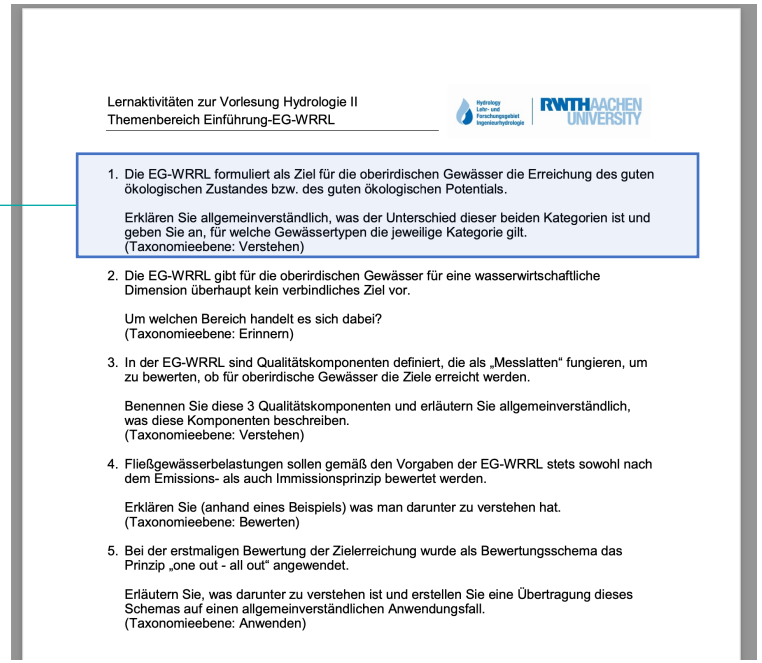
Example of think-back prompting



Beispiel zur Beantwortung von Fragen aus den Übungsblättern zwecks Modellevaluation

1.Die EG-WRRL formuliert als Ziel für die oberirdischen Gewässer die Erreichung des guten ökologischen Zustandes bzw. des guten ökologischen Potentials.

Erklären Sie allgemeinverständlich, was der Unterschied dieser beiden Kategorien ist und geben Sie an, für welche Gewässertypen die jeweilige Kategorie gilt.
(Taxonomieebene: Verstehen)



Lernaktivitäten zur Vorlesung Hydrologie II
Themenbereich Einführung-EG-WRRL

1. Die EG-WRRL formuliert als Ziel für die oberirdischen Gewässer die Erreichung des guten ökologischen Zustandes bzw. des guten ökologischen Potentials.

   Erklären Sie allgemeinverständlich, was der Unterschied dieser beiden Kategorien ist und geben Sie an, für welche Gewässertypen die jeweilige Kategorie gilt.
   (Taxonomieebene: Verstehen)

2. Die EG-WRRL gibt für die oberirdischen Gewässer für eine wasserwirtschaftliche Dimension überhaupt kein verbindliches Ziel vor.

   Um welchen Bereich handelt es sich dabei?
   (Taxonomieebene: Erinnern)

3. In der EG-WRRL sind Qualitätskomponenten definiert, die als „Messlatten" fungieren, um zu bewerten, ob für oberirdische Gewässer die Ziele erreicht werden.

   Benennen Sie diese 3 Qualitätskomponenten und erläutern Sie allgemeinverständlich, was diese Komponenten beschreiben.
   (Taxonomieebene: Verstehen)

4. Fließgewässerbelastungen sollen gemäß den Vorgaben der EG-WRRL stets sowohl nach dem Emissions- als auch Immissionsprinzip bewertet werden.

   Erklären Sie (anhand eines Beispiels) was man darunter zu verstehen hat.
   (Taxonomieebene: Bewerten)

5. Bei der erstmaligen Bewertung der Zielerreichung wurde als Bewertungsschema das Prinzip „one out - all out" angewendet.

   Erläutern Sie, was darunter zu verstehen ist und erstellen Sie eine Übertragung dieses Schemas auf einen allgemeinverständlichen Anwendungsfall.
   (Taxonomieebene: Anwenden)

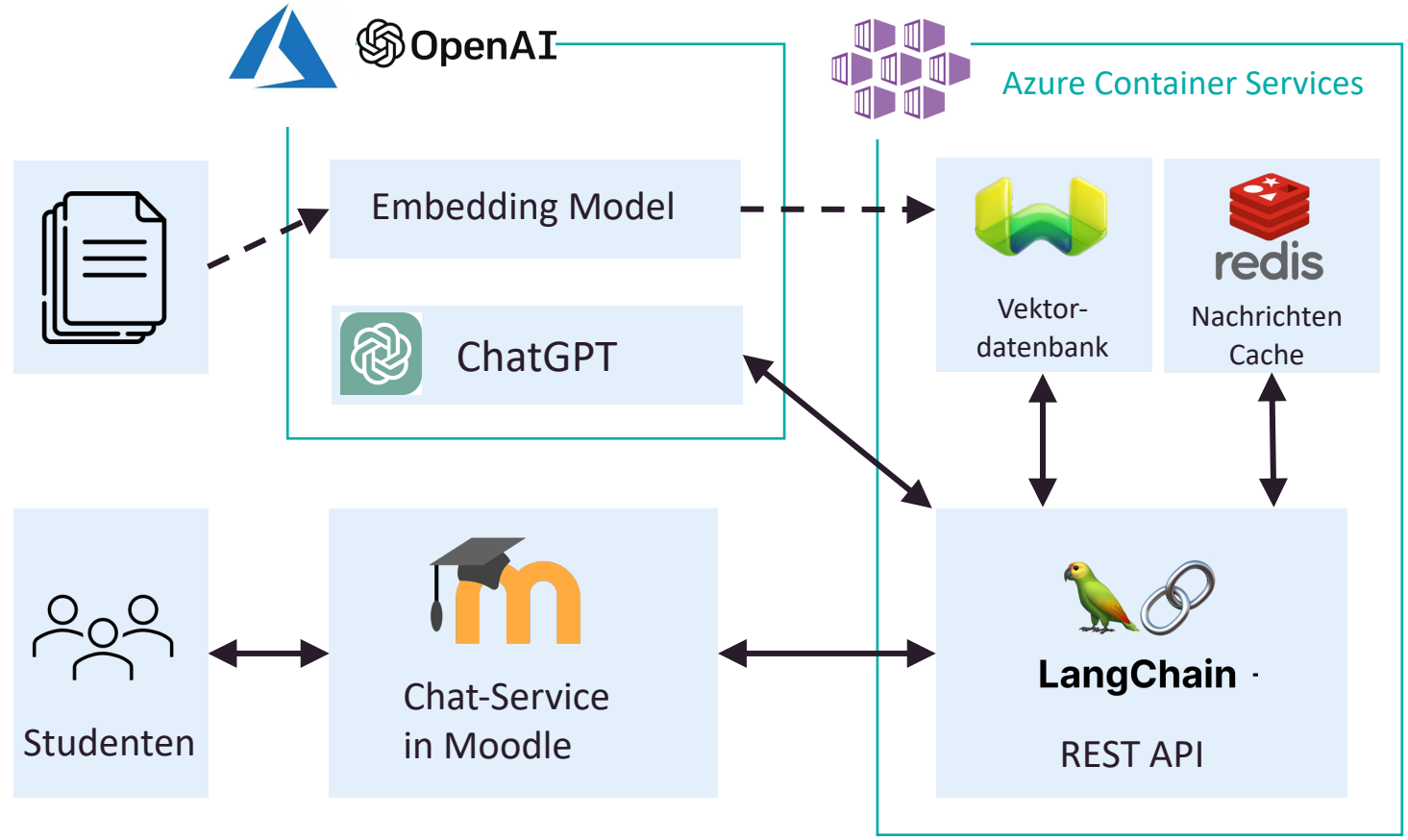
LangChain

```
USER_QA_PROMPT_TEMPLATE = """\
In the CONTEXT you will find the most relevant
information on the question from the course. \
Only use those information in the context.
Always mention the lecture and the page of the
used documents at the end of your answer. \
For example: "Hydrologie ist die Wissenschaft,
die sich mit dem Wasser auf der Erde befasst. \
(Lecture Hydrologie-II-1-7-Grundlagen, Seite 6)"

Additionally use the chat history between the
student (Human) and yourself (Chatbot) which \
provides context for conversation. \
If you can not help the user because you don't
have enough context, return "Das weiß ich auch
nicht".
Keep the answers short, focused and in the
language of the user.

<<CHAT HISTORY CONTEXT>>
{chat_history_context}

<< CONTEXT >>
{context}

<<CHAT HISTORY>>
{chat_history}
Human: {question}
"""
```

Der gute ökologische Zustand bezieht sich auf natürliche Gewässer, während das gute ökologische Potential für erheblich veränderte oder künstliche Gewässer gilt. Der Unterschied zwischen den beiden Kategorien liegt darin, dass der gute ökologische Zustand ein höheres Schutzniveau für die aquatische Umwelt darstellt als das gute ökologische Potential. Der gute ökologische Zustand bedeutet, dass das Gewässer in einem natürlichen Zustand ist und eine hohe biologische Vielfalt aufweist. Das gute ökologische Potential hingegen bedeutet, dass das Gewässer aufgrund menschlicher Eingriffe nicht mehr in einem natürlichen Zustand ist, aber dennoch ein hohes Potenzial für eine gute ökologische Qualität hat.
(Lecture: Hydrologie-II-1-1-Grundlagen-EG-WRRL, Page: 6)

Beispiel des Technologie-Stacks rund um den RAG Use-Case

Service-Architektur (am Bsp. Moodle-Integration)

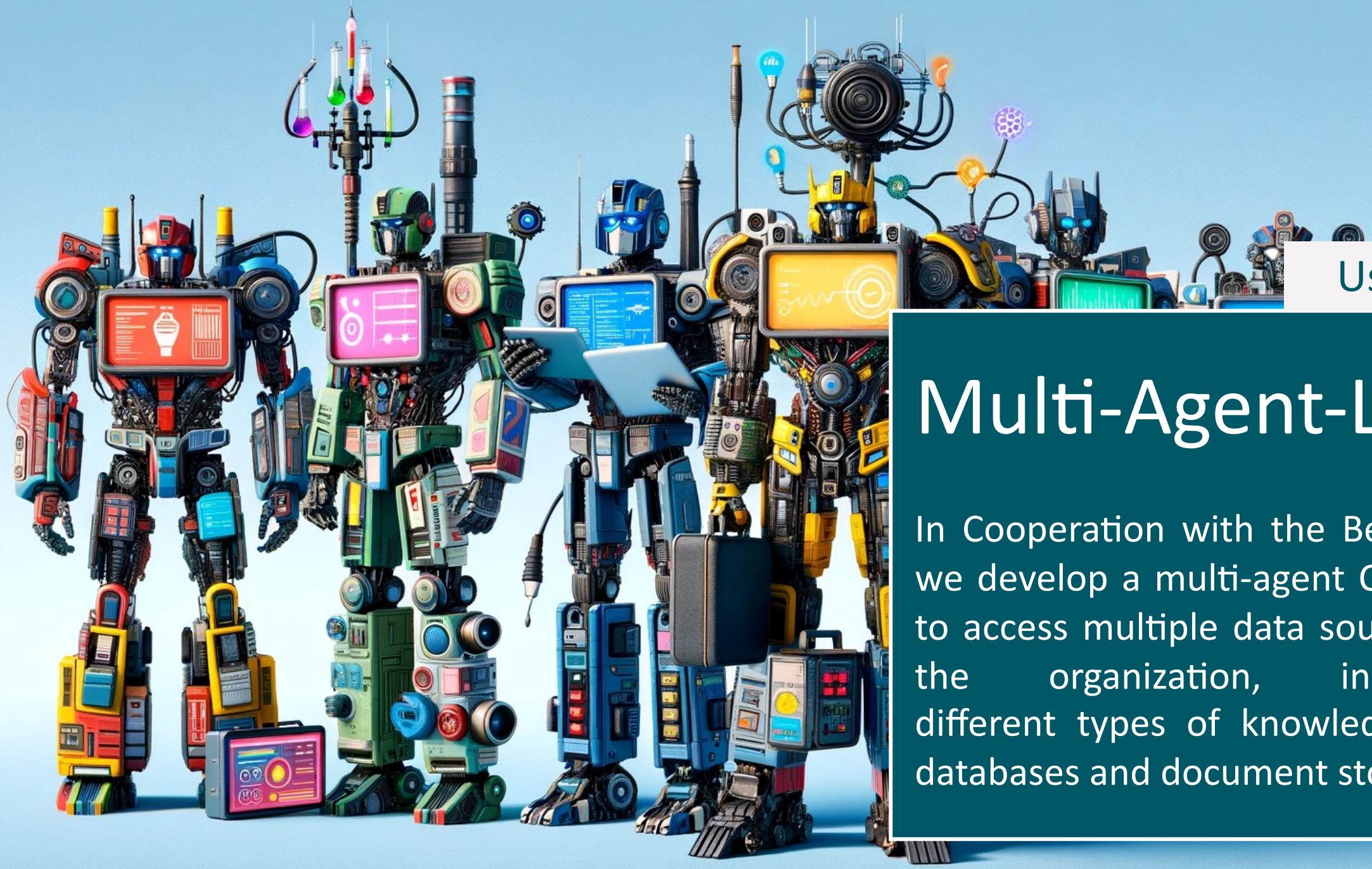


Daten import

Nutzung

Erklärung

- Vorlesungsunterlagen werden per OpenAI-Embedding[1] Modell in **Vektor-Repräsentationen umgewandelt** und in einer Vektor-Datenbank gespeichert.
- Studenten kommunizieren mittels Moodle mit unserem REST-Service.
- Dieser extrahiert mittels **Dokument-Retriever** die relevante Vorlesungs-informationen sowie vorausgegangen Chat-Nachrichten.
- Mittels LangChain werden diese Informationen zusammengeführt und durch die **ChatGPT API** beantwortet.



Usecase 2

Multi-Agent-LLMs

In Cooperation with the Bertelsmann, we develop a multi-agent Chat system to access multiple data sources across the organization, incorporating different types of knowledge graphs, databases and document stores.

From Large Language Models to Agents: What are the differences?

Large-Language-Models

- Trained to do texte generation
- Does not have any history (stateless)
- Can recommend call a function (i.e. to execute some Cypher-Query)

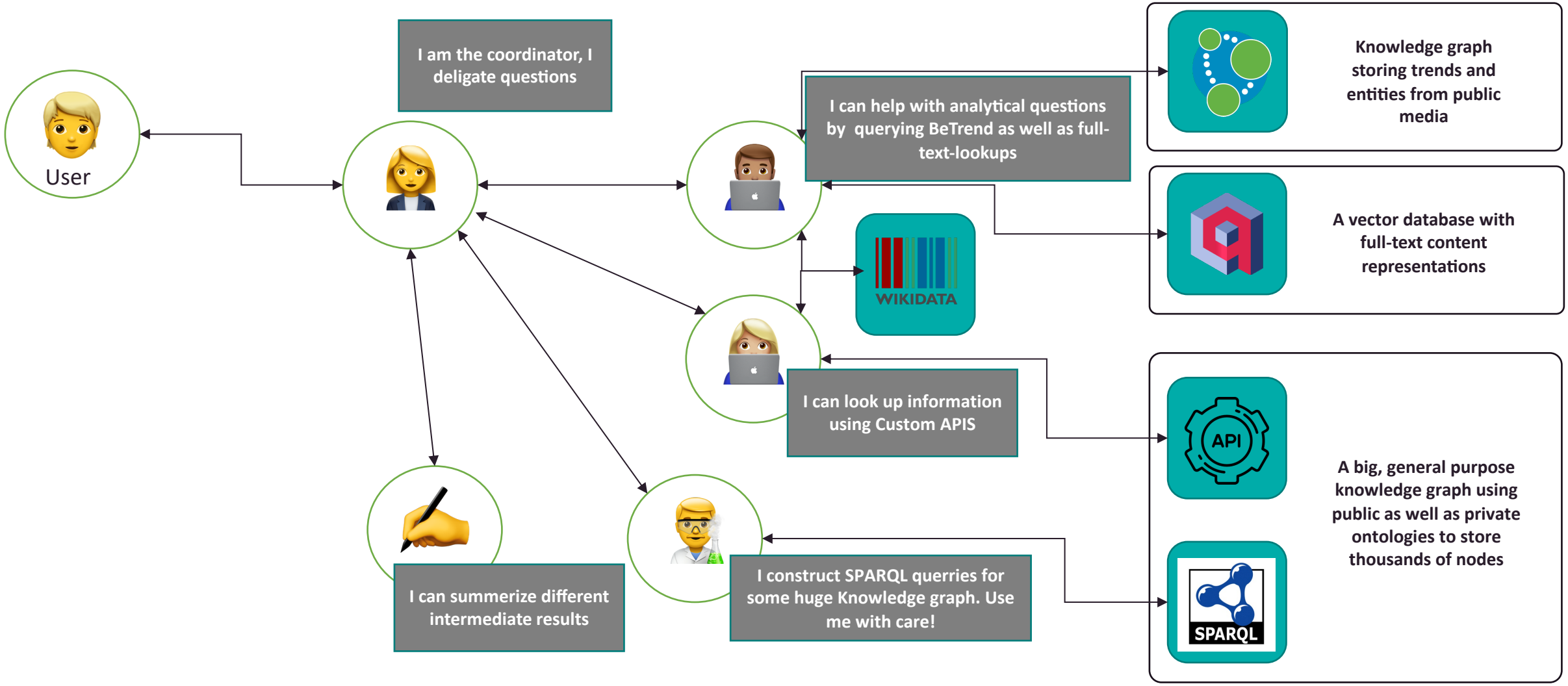
Agents

- Leverage power of LLMs to implement custom business logic
- Remember a full conversation (statefull)
- Can execute functions (i.e. run a cypher query and return the result

Better together: Improving LLMs using specialization

- Idea: Having **dedicated tasks for different agents** and allow them to collaborate
- Benefits of multiple agents:
  - Use different models and hyper-paremters for each agent
  - Use specific system prompts and tools for different roles
  - Store only relevant parts of message-history for each agent
- Challenges having a multi-agent-system:
  - Communication between agents requires an additional group chat manager (might make errors and additional LLM calls)
  - More complex technical setup
- Realization:
  - We switched from self-written agent class to Microsofts AutoGen Framework, that aims to make multi-agent conversations seamless

In the final system, a coordinator can delegate user questions to different specialized agents that collaborate to fulfill the task



Usecase 3

Personalization

LLMs can rewrite text in styles ranging form prose to legal text, from Shakespeare to Snoop Dog . Can we improve the interest in texts by personalizing them to the reader?

How to build a conference agenda recommender

scieneers
DRIVEN BY DATA

IT-Tage 2023

Crawl content from webpage

LangChain

Store summerized text and vector-representation

Retrieve candidates using vector-search

Pick top candidate and create individual recommendation text

Streamlit

More Details on the corresponding blog-post

scieneers
DRIVEN BY DATA

Looking for a partner

- That reaches thousands of users, i.e.
  - Visitors on a webpage
  - Readers of a newsletter
  - Receivers of advertisment-offers
- Has some information about those users
- Wants to test with us, if user acceptance can be improved using automatic personalization techniques

Visit the IT-Tage Program Advisor

Visit the associated blog-post:

Q&A

Contact us, if you are interested

Nico Kreiling

Alina Bickel